

# Tongda Xu

[x.tongda@nyu.edu](mailto:x.tongda@nyu.edu) | 185-1020-1763 | Github: <https://github.com/tongdaxu>

Apt 1507, 62 Xinwenhua St, Xicheng Dist, Beijing

I am a PhD candidate in Tsinghua University, advised by Dr. Ya-Qin Zhang. My research interests include image tokenization, compression and diffusion models. I have published 5 papers as first author in NIPS, ICML and ICLR. I am a contributor to Huggingface/diffusers and Zhipu/GLM-Image.

## EDUCATION

### Tsinghua University

*Doctor of Philosophy in Computer Science & Technology, Advised by Dr. Ya-Qin Zhang*

Beijing, China

Sep 2023 - Present

- **Visiting PhD in University of Cambridge:** Advised by Dr. José Miguel Hernández-Lobato.
- **Relevant Coursework:** Applied Stochastic Process, Information Theory, Advanced Theoretical Computer Science

### New York University

*Master of Science in Digital Media, GPA: 3.85/4.0, Advised by Dr. Wang Yao*

New York, US

Sep 2018 - Dec 2020

- **Awards:** Graduate Fellowship, Graduate Research Fellowship
- **Relevant Coursework:** Image and Video Processing, Machine Learning, Probability and Stochastic Process, Optimization and Numerical Linear Algebra, Data Structure and Algorithm, Computer System Architecture

### Tsinghua University

*Bachelor of Architecture, GPA: 3.82/4.0, Advised by Dr. Huang Weixin*

Beijing, China

Aug 2013 - May 2018

- **Awards:** National Fellowship, Tsinghua-Nitianzeng Fellowship, Academic Excellence Fellowship ×2

## EXPERIENCE

### Zhipu AI

*Research Intern (Text-to-Image Foundation Model)*

Beijing, China

Feb 2025 - now

- **GLM-Image:** Training the feature aligned VAE of GLM-Image diffusion model.
- **Image Tokenization using Gaussian VAE:** Building a image tokenizer from continous Gaussian VAE, submitted to ICLR 26.
- **Cogview 4 model Distillation:** Distillation of 10B text-to-image diffusion model.

### NVIDIA Corporation

*AI4Science Intern, Advised by Zhixiang Dai*

Beijing, China

Nov 2024 - Feb 2025

- **Estimating Sample Uncertainty in Diffusion based AI4Science model:** Fast sample variance estimation for diffusion based surface solar radiation downwards prediction. Project presented in GTC 2025.

### Microsoft Research Asia (MSRA)

*Research Intern (Media Computing Group), Excellent Intern, Advised by Dr. Li Bin*

Beijing, China

Feb 2024 - Nov 2024

- **Diffusion based Screen Content Compression:** Construct a screen image codec using diffusion model, published in CVPR 25 [c11].

### Institute for AI Industry Research (AIR), Tsinghua University

*Research Assistant (Neural Codec), Advised by Dr. Wang Yan*

Beijing, China

June 2022 - Present

- **Bit-allocation using Optimization :** Prove the equivalence between optimal bit allocation for video codec via semi-amortized variational inference. Published in ICML 2023.

### Sensetime Research

*Researcher (Codec Group), Advised by Dr. Wang Yan*

Beijing, China

Jan 2021 - June 2022

- **Semi-amortized Inference for Variable Bitrate Image Compression:** Propose an approach to achieves variable bitrate model and perception-distortion trade-off in neural image compression using semi-amortized variational inference. Published in NIPS 2022 [c6].
- **Importance Weighted Neural Image Compression:** We propose an approach for neural image compression training with tighter ELBO. Published in NIPS 2022 [c5].
- **Spatial Moment Pooling for Blind Image Quality Assessment:** Propose to extent spatial average pooling into spatial moment pooling by incorporating higher order moments. Published in ICIP 2022 [c4].
- **Compressed Domain Contour Flow for Fast Video Salient Detection:** Propose a novel contour-flow approach for P/B frame salient object detection by wrapping the I frame contours with motion vectors encoded in bitstream. The boundary problem of feature-flow approaches is overcame by wrapping contours directly. Achieve 400% speed up with 3% f-measure loss compared with SOTA.

### Microsoft Research Asia (MSRA)

*Research Intern (Media Computing Group), Excellent Intern, Advised by Dr. Li Bin*

Beijing, China

Jan 2020 - Dec 2020

- **Joint Rate Control for Real-time Communication with Reinforcement Learning:** Propose and implement a deep reinforcement learning based framework for joint rate control of multiple sources in real-time communication. Outperform WebRTC + VP8 in screen and video sharing scenario by 0.5db in PSNR.

- **Fluid Screen Sharing:** Propose and implement a mixed-integer linear programming based algorithm for screen content re-layout. Facilitate real-time relaying out for screen sharing across different devices.
- **Error Recovery for Video Communication:** Dive in, implement and evaluate core video error recovery algorithm including Hybrid Type I, WebRTC and Microsoft in-house RTC.

### Video Lab, New York University

Research Intern (Image Processing), Advised by Dr. Wang Yao

New York, US  
May 2019 - Aug 2019

- **Optimal Feature Subset Selection for MRI:** Propose a novel improvement of A\* heuristic search for feature selection, combining the idea of A\* graph search and genetic algorithm, improved state of the art of problem by 0.2 in r2 score for regression.
- **Image Cascade Network for Large 3D Medical Image Segmentation:** Propose and implement an end-to-end auto-context based multi-level network structure enabling semantic segmentation for 3D image as large as 256<sup>3</sup>. Achieve dice coefficient 0.99 for background, 0.92 for mouse embryo and 0.85 for brain vehicle.

### Aedas Ltd (HK Office)

Architecture Intern, Advised by Christine Lam and David Clayton

Hong Kong, China  
May 2017 - Aug 2017

- **Computer-Aided Design Toolkit:** Optimize working pipeline and accelerate design, calculation and analysis process from feasibility study to construction for 6 complex buildings in China. Compile a computational toolkit with C# in Grasshopper, facilitating building codes calculation and parametric curtain wall design.

### X-Studio, Tsinghua University

Research Intern (Human-Computer Interaction), Advised by Dr. Mi Haipeng

Beijing, China  
May 2015 - Aug 2015

- **Tangible Circuit Training:** Aid design and implement of a table based tangible user interface application simulating circuits and electric components. Aid graphic and interface programming in Java.

## TEACHING

### DM-UY 1143 - Ideation & Prototyping, New York University

Teaching Assistant to Ideation and Prototyping of Benedetta Piantella

New York, US  
Sep 2018 - Jan 2019

### Architecture Design Studio 6 - Parametric Design, Tsinghua University

Teaching Assistant to Parametric Design studio of Dr. Weixin Huang

New York, US  
May 2015 - Aug 2015

## SELECTED PUBLICATIONS

- [c11]: **Xu, T.**, Li, J., Li, B., Wang, Y., Zhang, Y. Q., Lu, Y. Versatile Perceptual Image Compression with Diffusion Rendering. CVPR 25.
- [c10]: **Xu, T.**, Cai, X., Zhang, X., Ge, X., He, D., Sun, M., ... Wang, Y. Rethinking Diffusion Posterior Sampling: From Conditional Score Estimator to Maximizing a Posterior. ICLR 25.
- [c9]: Zhang, X., Ge, X., **Xu, T.**, He, D., Wang, Y., Qin, H., ..., Zhang, J. Gaussianimage: 1000 fps image representation and compression by 2d gaussian splatting. ECCV 25.
- [c8]: **Xu, T.**, Zhu, Z., He, D., Li, Y., Guo, L., Wang, Y., ... Zhang, Y. Q. Idempotence and Perceptual Image Compression. ICLR 24 spotlight.
- [c7]: Ge, X., Luo, J., Zhang, X., **Xu, T.**, Lu, G., He, D., ... Qin, H. Task-aware encoder control for deep video compression. CVPR 24.
- [c6]: Zhang, X., Yang, R., He, D., Ge, X., **Xu, T.**, Wang, Y., ..., Zhang, J. Boosting neural representations for videos with a conditional decoder. CVPR 24 highlight.
- [c5]: Li, Y., **Xu, T.**, Wang, Y., Liu, J., Zhang, Y. Q. Idempotent Learned Image Compression with Right-Inverse. NIPS 24.
- [c4]: **Xu, T.\***, Gao, H.\*, Gao, C., Pi, J., Li, Y., Wang, Y., ... & Wang, Y. Bit allocation using optimization. ICML 23 (\*equal contribution)
- [c3]: Gao, C.\*, **Xu, T.\***, He, D., Wang, Y., Qin, H. Flexible Neural Image Compression via Code Editing. NIPS 22 (\*equal contribution)
- [c2]: **Xu, T.**, Wang, Y., He, D., Gao, C., Gao, H., Liu, K. Qin, H. Multiple-sample Neural Image Compression. NIPS 22.
- [c2]: He, D., Yang, Z., Yu, H., **Xu, T.**, Luo, J., Chen, Y., ..., Wang, Y. Po-elic: Perception-oriented efficient learned image coding. CVPRW-CLIC 2022 1st place.
- [c1]: **Xu, T.**, Wang, D., & You, X. Mindgame: Mediating People's EEG Alpha Band Power through Reinforcement Learning. UIST 18 (pp. 5-6).

## COMMUNITY SERVICE

- **Program Committee / Reviewer:** NIPS 23-25, ICML 25, ICLR 24-26, CVPR 23-26, ICCV & ECCV 23-25, AAAI 25-26, MM 25, AISTATS 25, TNNLS, TITS, TGRS, RA-L, etc.

## PATENTS

- **Granted:** CN113660531B, CN113612999B, CN115695800B